

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平6-308994

(43)公開日 平成6年(1994)11月4日

(51)Int.Cl. <sup>5</sup>	識別記号	府内整理番号	F I	技術表示箇所
G 10 L 3/00	5 3 1 D	9379-5H		
	F	9379-5H		
	5 3 5	9379-5H		
	5 6 1 G	9379-5H		

審査請求 未請求 請求項の数 2 O L (全 8 頁)

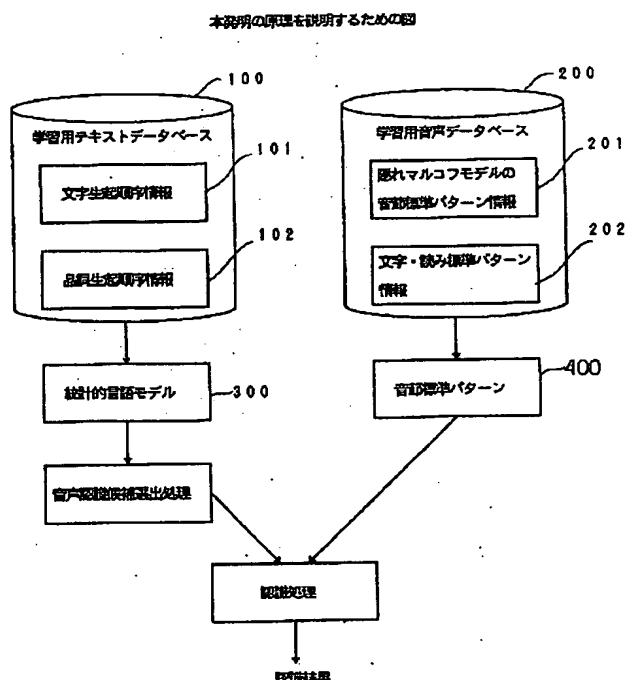
(21)出願番号 特願平5-99694	(71)出願人 日本電信電話株式会社 東京都千代田区内幸町一丁目1番6号
(22)出願日 平成5年(1993)4月26日	(72)発明者 山田 智一 東京都千代田区内幸町1丁目1番6号 日本電信電話株式会社内
特許法第30条第1項適用申請有り 1993年1月18日、社団法人電子情報通信学会発行の「電子情報通信学会技術研究報告V o l. 92, N o. 409」に発表	(72)発明者 松永 昭一 東京都千代田区内幸町1丁目1番6号 日本電信電話株式会社内
	(72)発明者 鹿野 清宏 東京都千代田区内幸町1丁目1番6号 日本電信電話株式会社内
	(74)代理人 弁理士 伊東 忠彦

(54)【発明の名称】 日本語音声認識方法

(57)【要約】

【目的】 本発明の目的是、文法的に誤った候補を統計的に削除し、単純な文字の組み合わせによる候補が多数生成されることを防ぐことにより処理時間を短縮し、変換性能を向上させることができる日本語音声認識方法を提供することである。

【構成】 本発明は、統計的言語モデルとして、学習用テキストデータベース100から作成された文字の生起順序情報101、及び品詞の生起順序情報102に関する統計的言語モデルを用い、標準パターンとして、学習用音声データベース200の音節標準パターン201及び文字の読みに対する音節標準パターン202を用いる。



## 【特許請求の範囲】

【請求項1】 入力音声を特徴パラメータの時系列とし、学習用テキストデータベースより作成された生起順序に関する統計的言語モデルを用いて、該入力音声の特徴パラメータ時系列について、複数の音声認識候補を選び出し、これらの各音声認識候補について、隠れマルコフモデルの標準パターンと該入力音声の特徴パラメータ時系列とをそれぞれ照合して、生起の尤度と類似の尤度との総合尤度の高い候補を認識結果とする日本語音声認識方法において、  
該統計的言語モデルとして、該学習用テキストデータベースから作成された、文字の生起順序、及び品詞の生起順序に関する統計的言語モデルを用いることを特徴とする日本語音声認識方法。

【請求項2】 入力音声を特徴パラメータの時系列とし、学習用テキストデータベースより作成された生起順序に関する統計的言語モデルを用いて、該入力音声の特徴パラメータ時系列について、複数の音声認識候補を選び出し、これらの各音声認識候補について、隠れマルコフモデルの標準パターンと該入力音声の特徴パラメータ時系列とをそれぞれ照合して、生起の尤度と類似の尤度との総合尤度の高い候補を認識結果とする日本語音声認識方法において、  
該標準パターンとして、音節標準パターン及び文字の読みに対する音節標準パターンを用いることを特徴とする日本語音声認識方法。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】 本発明は、日本語音声認識方法に係り、特に、音韻を認識する方法としてパターンマッチングに基づくものと、特徴抽出に基づくもののうち、確率的に状態遷移を行い、その状態遷移の際のある確率で記号を出力する隠れマルコフモデル（例えば、中川聖一「確立モデルによる音声認識」電子情報通信学会編（1988））と、統計的言語モデル（例えば、L.R. Bah1他“A Statistical Approach to Continuous Speech Recognition” IEEE Trans. on PAMI(1983)）とを用いた日本語音声認識方法に関する。

## 【0002】

【従来の技術】 図4は、従来の日本語音声認識システムを説明するための図である。

【0003】 同図に示すシステムは、仮名・漢字生起順序に関する統計的言語モデル情報11を有する学習用テキストデータベース10、学習用テキストデータベース10の統計的言語モデル情報11を予め記憶しておく統計的言語モデルメモリ50、隠れマルコフモデルの音素標準パターン情報21を有する学習用音声データベース20、学習用音声データベースの音素標準パターン情報21を予め記憶しておく文字・読みに対応する音素系列メモリ60、入力音声に対して既に認識処理が済んでい

る直前の音素から音素候補を選出する音素候補選出部30、学習用テキストデータベース10の標準モデルと入力音声を照合し、学習用音声データベース20の音素標準パターンの情報により総合的尤度の最も高い候補の仮名・漢字文字を認識し、出力する認識部40より構成される。

【0004】 次に、従来のシステムの動作を説明する。従来の隠れマルコフモデル及び統計的言語モデルを用いた音声認識で入力音声を仮名・漢字系列に変換する方法

10 として、学習用テキストデータベース10より、仮名・漢字の生起順序に関する統計的言語モデル11を、学習用音声データベース20より隠れマルコフモデルの音素標準パターン21を、各々統計的言語モデルメモリ50、文字読みに対応する音素系列メモリ60に予め作成しておき、音声候補選出部30で入力音声に対して、統計的言語モデル11と、文字の読みに対応する音素系列の情報を用いて、既に認識した直前の複数の音素から、次に生起する確率の高い複数の音素候補を選出し、これら選出した音素候補のそれぞれについて、認識部40で20 その音素標準パターンと入力音声とを照合して、統計的言語モデルによる生起尤度と、隠れマルコフモデルによる標準パターンとの類似尤度との総合的尤度の最も高い候補の仮名・漢字文字を認識結果として出力することが提案されている。

## 【0005】

【発明が解決しようとする課題】 しかしながら、上記従来の変換方法は、統計的言語モデルが仮名・漢字の文字面だけを扱い、また、音素系列から仮名・漢字文字への変換文字への変換情報として、各仮名・漢字文字の読み30 に対応する音韻系列の情報が与えられているだけでないで、単純な文字の組み合わせによる候補が多数生成されるという問題がある。

【0006】 本発明は、上記の点に鑑みなされたもので、上記従来の問題点を解決し、文法的に正しい候補を統計的に選択することで、文法的に誤った候補を統計的に削除し、単純な文字の組み合わせによる候補が多数生成されることを防ぐことにより処理時間を短縮し、変換性能を向上させることができる日本語音声認識方法を提供することを目的とする。

## 【0007】

【課題を解決するための手段】 図1は、本発明の原理を説明するための図である。

【0008】 本発明は、入力音声を特徴パラメータの時系列とし、学習用テキストデータベース100より作成された生起順序に関する統計的言語モデル101を用いて、入力音声の特徴パラメータ時系列について、複数の音声認識候補を選び出し、これら各音声認識候補について、学習用音声データベース200の隠れマルコフモデルの標準パターンと入力音声の特徴パラメータと入力音50 声の特徴パラメータ時系列とをそれぞれ照合して、生起

の尤度との総合尤度の高い候補を認識結果とする日本語音声認識方法において、統計的言語モデルとして、学習用テキストデータベース100から作成された文字の生起順序情報101、及び品詞の生起順序情報102に関する統計的言語モデルを用いる。

【0009】また、本発明は、入力音声を特徴パラメータの時系列とし、学習用テキストデータベース100より作成された生起順序に関する統計的言語モデルを用いて、入力音声の特徴パラメータ時系列について、複数の音声認識候補を選出し、これら各音声認識候補について、学習用音声データベース200の隠れマルコフモデルの標準パターンと入力音声の特徴パラメータと入力音声の特徴パラメータ時系列とをそれぞれ照合して、生起の尤度との総合尤度の高い候補を認識結果とする日本語音声認識方法において、標準パターンとして、学習用音声データベース200の音節標準パターン201及び文字の読みに対する音節標準パターン202を用いる。

#### 【0010】

【作用】本発明は、学習用テキストデータベースの情報として仮名・漢字の生起順序に加え、品詞の生起順序を考慮し、学習用音声データベースの情報として、隠れマルコフモデルの音素標準パターン情報（音節標準パターン）に加え、文字の読み標準パターン情報を考慮することにより、単純な文字の組み合わせにより生成される候補に比べ、文法的に正しい候補が統計的に選択され、生成される候補が絞り込まれるので、入力された音声から仮名・漢字系列へ変換するのに要する処理時間を短縮し、変換性能を向上させることができる。

#### 【0011】

【実施例】以下、図面と共に本発明の実施例を説明する。

【0012】図1は本発明の一実施例のプロック図である。

【0013】同図に示すシステムは音声信号を入力する音声信号入力端子1、特徴パラメータを抽出する特徴抽出部2、文字候補を選出し、特徴パラメータとの類似尤度を求める認識部3、音節連鎖標準パターンを記憶する標準パターンメモリ4、文字品詞統計的言語モデルを記憶する文字・品詞統計的言語モデルメモリ5、及び認識部3からの認識結果を出力する認識結果出力部6より構成される。

【0014】まず、最初に予め、学習用テキストデータベース100の生起順序情報101及び品詞生起順序情報102により統計的言語モデルを記憶する標準パターンメモリ4と、学習用音声データベース200の隠れマルコフモデルの音声標準パターン情報201、文字読み

$$P(C_i \mid C_{i-2}, C_{i-1})$$

となる。このデータを得るためのテキスト及び品詞列が十分に用意できない場合は、

標準パターン情報202により文字・品詞統計的言語モデルを読みだし、文字・品詞統計的言語モデルメモリ5を生成する。

【0015】入力端子1から入力された音声は、特徴抽出部2でデジタル信号に変換され、更にLPCケプストラム分析された後、1フレーム（例えば、10ミリ秒）毎に特徴パラメータに変換される。この特徴パラメータは、例えば、LPCケプストラム係数である。

【0016】学習用音声データベース200より、特徴ベクトルと同一形式で、隠れマルコフモデルの音節標準パターン及び漢字の読みに対する音節連鎖標準パターン等の標準パターンを予め作り、標準パターンメモリ4に記憶しておく、また、学習用テキストデータベース100より文字（仮名・漢字）及び品詞の生起順序に関する統計的言語モデルを作成し、文字・品詞統計的言語モデルメモリ5に記憶してある。

【0017】認識部3は、文字・品詞統計的言語モデルメモリ5に記憶されている文字・品詞統計的言語モデルを用いて選出した複数の文字候補について、その文字候補の読みを表す標準パターンを標準パターンメモリ4から読み出し、入力音声の特徴パラメータとの類似尤度をそれぞれ求める。

【0018】認識結果出力部6は、認識部3の認識結果に基づいてディスプレイ等に出力する。

【0019】図3は本発明の一実施例の認識処理を説明するための図であり、同図は、認識部3においてi番目の認識のための候補文字を選出し、これより認識結果を出力する場合を示す。つまり、例えば、入力音声に対するi番目の文字を認識するには、文字・品詞統計的言語モデルから仮名・漢字と品詞の生起順序に関する条件つき確率を用いて、(i-2)番目と(i-1)番目との各文字の認識結果と、(j-1)番目と(j-2)番目との各単語の品詞の認識結果と、次に現れると仮定されるj番目の単語の品詞とを基に、i番目に出現すると予測される尤度が高い文字の複数を候補文字k<sub>1</sub>～k<sub>n</sub>として選出す。

【0020】ここで、i番目に出現すると予測される文字は、j番目の単語の一部である。j番目の単語としては、存在する全ての品詞を仮定してもよいし、i番目に出現すると予測される文字により仮定できる品詞を限定してもよい。i-2, i-1, i番目に現れる文字をc<sub>i-2</sub>, c<sub>i-1</sub>, c<sub>i</sub>とし、j-2, j-1, j番目に現れる単語の品詞をs<sub>j-2</sub>, s<sub>j-1</sub>, s<sub>j</sub>とすると、前述の条件つき確率は、

#### 【0021】

#### 【数1】

$$S_{i-2}, S_{i-1}, S_i)$$

#### 【0022】

#### 【数2】

$$P(C_i | C_{i-2}, C_{i-1}, C_i) \times P(S_i | S_{i-2}, S_{i-1})$$

などを用いて同様の効果を得ることも可能である。また、 $i-2$ ,  $i-1$ ,  $i$ 番目に現れる  $c_{i-2}$ ,  $c_{i-1}$ ,  $c_i$  としては、単純な文字そのものではなく、予め読みの情報を付与した文字コードを用いても差し支えない。

【0023】次にこれら選出された各候補文字に対し、仮定する品詞の数に応じて別々の候補を作成する。

【0024】例えば、図2に示すように、文字  $k_1$  に対し、品詞  $h_1$ ,  $h_2$ ,  $h_3$  を仮定する場合、その組み合わせによる3つの候補が作成される。そしてこれらの候補の各々について、標準パターンメモリ4から文字の読みに対する標準パターンを読みだし、 $i$ 番目の入力音声との尤度を求め、文字・品詞統計的言語モデルメモリ5から読みだした文字・品詞統計的言語モデルに基づくその候補文字の  $i$  番目に生起する尤度との和を総合尤度とし、この総合尤度の最も高い候補、例えば、 $k_1$ ,  $h_2$  が最も総合尤度が高ければ、文字  $k_1$  を  $i$  番目の認識結果文字として認識結果出力部6へ出力する。

【0025】ある文字が出力されたとき、総合尤度が最も高かった候補が次の文字を出力した段階でも総合尤度が最も高くなるとは限らない。そこで、総合尤度の高い上記B個の候補を残し（これをピーム幅がBであると言う）、次の処理へ引き継ぐ。上位B個としたのは、すべての候補を保持するとメモリ量が増大し、それに伴い処理時間が長くなる等により困難であるためである。総合尤度の最も高い候補の変更に伴い、認識結果出力部6へ出力する候補文字も更新する。

【0026】この仮名・漢字候補選出及びその文字の品詞の仮定と、それらについての標準パターンとの照合と、その総合尤度から認識結果文字を尤度の高い順に仮名・漢字系列として出力する。

【0027】なお、入力音声中に、学習テキストデータベース中にはない漢字があると、これを認識することができない。この場合は、その認識できない文字（漢字）を、認識結果の仮名・漢字系列中に空白として出力するか、あるいは、音素、または仮名の生起順序に関する統計的言語モデルと、隠れマルコフモデルの音素または音節標準パターンとを掛け、データベースにない漢字は、音素系列または仮名系列として出力してもよい。

【0028】また、特徴抽出部2、認識部3、認識結果出力部6はそれぞれ専用または、兼用のマイクロプロセッサにより処理することができる。

#### 【0029】

【発明の効果】上述のように、本発明によれば、仮名・漢字の生起順序に関する統計的言語モデルと隠れマルコフモデルの音節標準パターン及び漢字の読みに対する音声連鎖標準パターンとを用いて、入力された音声から直接漢字仮名混じり系列を出力する場合において、文字だけでなく品詞の生起順序に関する情報も用いて予め統計的言語モデルを作成しておくことで、文法的に誤った候補を統計的に削除することができ、処理時間を短縮し、変換性能を向上させることができる。

#### 【図面の簡単な説明】

【図1】本発明の原理を説明するための図である。

【図2】本発明の一実施例のブロック図である。

【図3】本発明の一実施例の認識処理を説明するための図である。

【図4】従来の日本語音声認識システムを説明するための図である。

#### 20 【符号の説明】

1 音声信号入力端子

2 特徴抽出部

3 認識部

4 標準パターンメモリ

5 文字・品詞統計的言語モデル

6 認識結果出力部

10 1 学習用テキストデータベース

11 1 仮名・漢字の生起順序に関する統計的言語モデル情報

#### 30 20 学習用音声データベース

21 隠れマルコフモデルの音素・標準パターン情報

30 音素候補選出部

40 認識部

50 統計的言語モデルメモリ

60 文字の読みに対応する音素系列メモリ

100 学習用テキストデータベース

101 文字生起順序情報

102 品詞生起順序情報

200 学習用音声データベース

#### 40 201 隠れマルコフモデル音節標準パターン情報

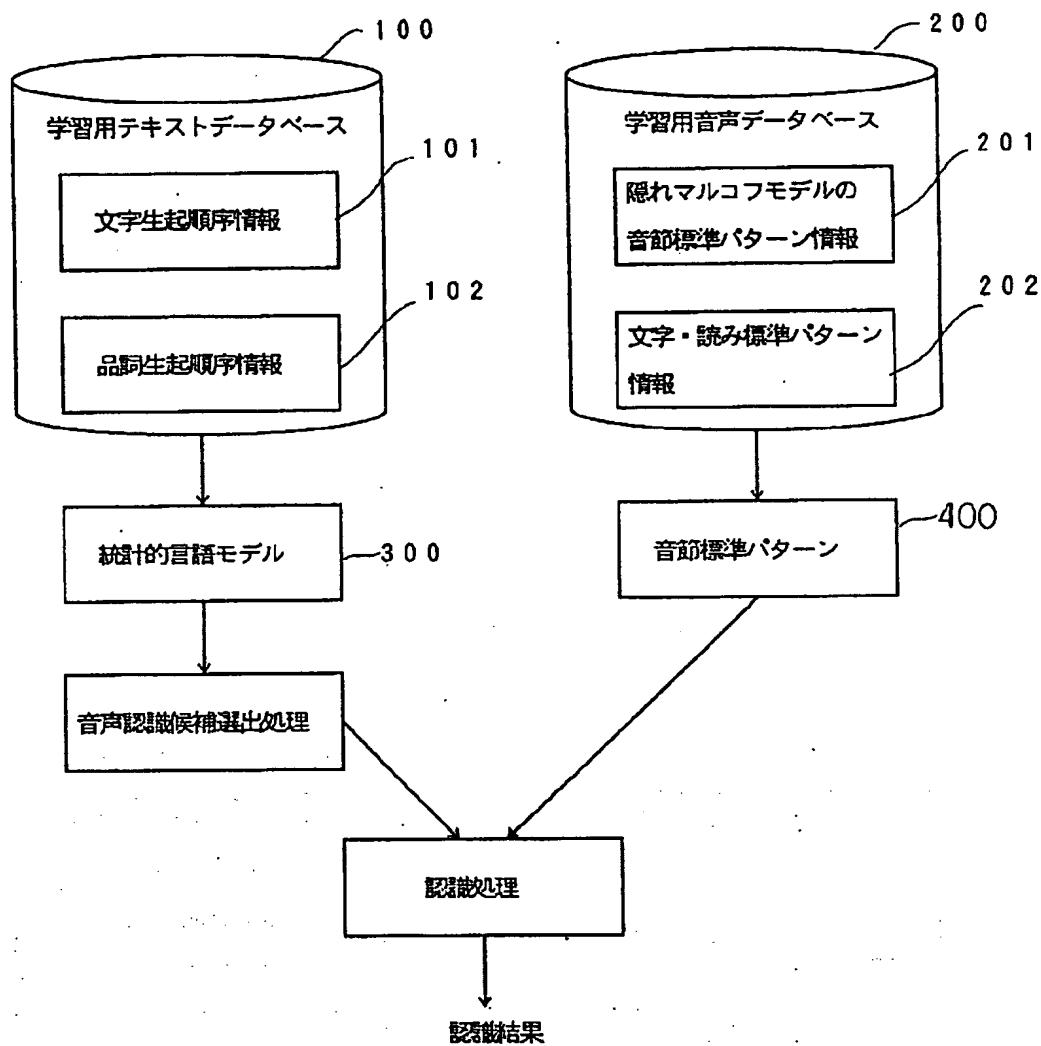
202 文字読み標準パターン情報

300 統計的言語モデル

400 音節標準パターン

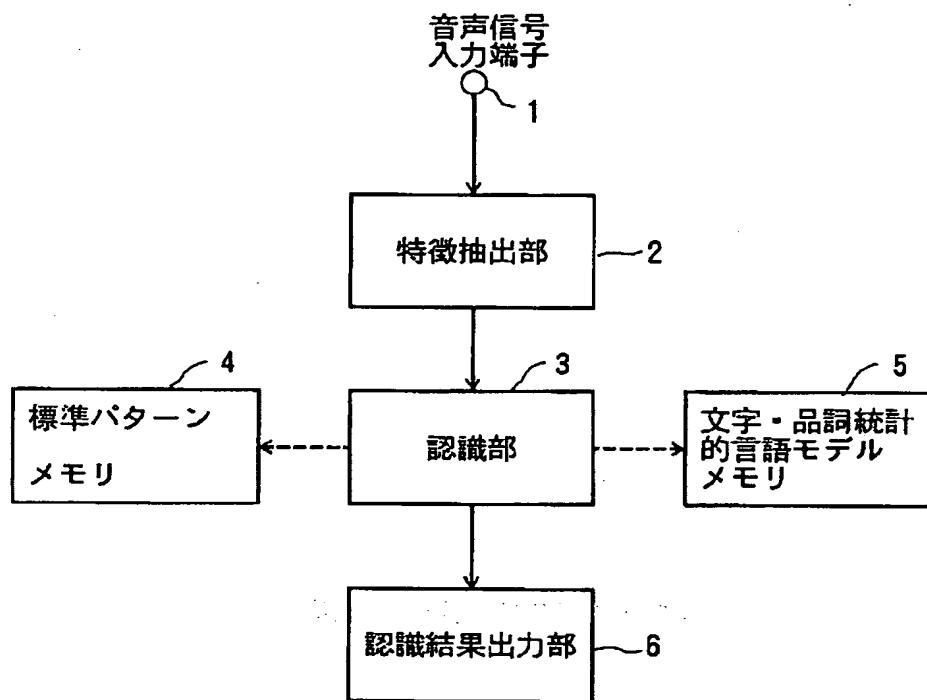
【図1】

本発明の原理を説明するための図



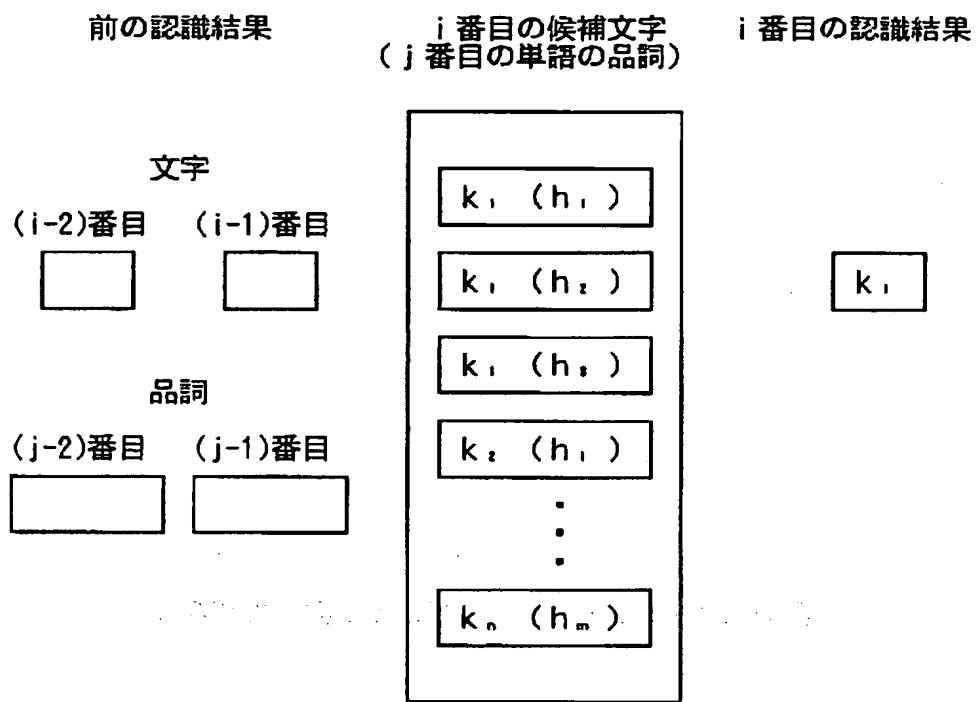
【図2】

## 本発明の一実施例のブロック図



【図3】

## 本発明の一実施例の認識処理を説明するための図



【図4】

従来の日本語音声認識システムを説明するための図

